



Citation	N. Reynders, W. Dehaene, (2014), A 210mV 5MHz Variation-Resilient Near-Threshold JPEG Encoder in 40nm CMOS Proceedings of IEEE International Solid-State Circuits Conference (ISSCC), 456-457.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1109/ISSCC.2014.6757511
Journal homepage	
Author contact	nele.reynders@esat.kuleuven.be + 32 (0)16 321104

(article begins on next page)



Nele Reynders, Wim Dehaene

KU Leuven, Leuven, Belgium

Operating circuits in the near-threshold region enables large energy savings. However, such circuits also pose many challenges, such as increased delay, unwanted leakage paths and high sensitivity to variations. Working in advanced nanometer CMOS technologies compromises the robustness of circuits even more due to the increased variability. Nonetheless, these technologies offer higher operating frequencies for ultra-low-voltage circuits. Transitioning to smaller technologies is attractive for digital near-threshold circuits, provided that the impact of the increased variability can be mitigated. Few prior works have considered the design of variation-resilient ultra-low-voltage circuits in CMOS technologies smaller than 65nm. The aim of this work is to design a large system that advances the state-of-the-art by not only reaching very low energy consumption, but also clock frequencies of tens of MHz, while providing high variation resilience. We present a full JPEG encoder fabricated in a 40nm CMOS technology, fully functional down to a supply voltage of 210mV. The JPEG encoder is able to operate at clock frequencies in a range from 5 to 275MHz for supplies from 210 to 550mV, and thus achieves very high ultra-low-voltage speed. At the minimum-energy point (MEP), the JPEG encoder consumes only 29.01pJ/pixel at an operating frequency of 41MHz ($V_{DD}=330mV$). The variation (σ/μ) of 26 dies at this point is only 9.4% for the frequency and 6.0% for the energy consumption.

The pipelined JPEG encoder, which is compliant with the baseline sequential mode of the JPEG image compression standard [1], consists of 3 main building blocks (Fig. 27.3.1). First, the discrete cosine transform (DCT) transforms blocks of 8x8 pixels to the frequency domain. Second, the quantization divides the transformed blocks by a specific quantization matrix, which determines the compression factor of the JPEG encoding. The quantization coefficients are stored in the quantization table. Third, the blocks are linearized in zigzag order to group similar frequencies. This data is then Huffman encoded according to the DC and AC Huffman tables.

The 2D-DCT (Fig. 27.3.2) is implemented as a sequence of two 1D-DCTs with a transpose matrix in between. The 1D-DCT is based on the algorithm proposed in [2] and consists of five 15b adder/subtractors and one 15b multiplier. The 2D-DCT output needs to be scaled by a scaling matrix afterwards, but this is performed without extra hardware, as it is incorporated in the quantization by scaling the quantization coefficients in advance. The quantization (Fig. 27.3.2) is computed by a multiplier that accesses the quantization table through a one-hot decoder. Fig. 27.3.1 also shows the different subblocks of the Huffman encoder. After the zigzagging, the DC component and the 63 AC components are encoded separately.

The JPEG encoder has a latch-based deeply pipelined architecture. The pipeline is controlled by non-overlapping clock signals, which are distributed to the entire chip by a clock tree. The timing block shown in Fig. 27.3.1 consists of the non-overlapping clock generator and the complete clock tree. The topology of the logic gates is critical for ultra-low-voltage designs, in terms of speed, energy consumption and variation resilience. Therefore, the topology of all logic gates used throughout this JPEG encoder is based on differential transmission gate (TG) logic extended with NMOS stacking [3]. A pipeline stage is at most 3 TG logic gates deep. This balances robustness and gate reliability on the one hand, and on the other hand, the averaging of timing variations obtained by cascading logic gates.

The quantization, DC and AC Huffman tables are implemented as register tables, where the desired content is serially shifted in at startup and the data can be accessed through a one-hot decoder in the case of quantization, and full address decoders in the case of the Huffman tables. The reason why the tables are not implemented as SRAM memories is because the energy consumption of SRAMs is not dominated by dynamic energy but rather, by standby leakage. Therefore, constructing the tables as sub-threshold SRAMs is not a good option. Furthermore, for the required numbers of bits in the design, the peripheral area and energy overhead would be too high, while the speed of the sub-threshold SRAM would be insufficient. Moreover, efficient SRAM requires ratioed logic,

which is undesirable in the ultra-low-voltage domain due to its variation sensitivity. The transpose and zigzag matrices consist of two 8x8 blocks of registers. They are serially read in the original order in the 1st block, then copied in parallel to the 2nd block from which they are serially read out in a different, desired order. The dense layout of the JPEG encoder is carried out using a software tool (the Datapath Generator [4]), except for the layout of the 3 tables, which was done manually because their regularity allowed an optimized structure. The active area of the JPEG encoder is 0.557mm² (Fig. 27.3.7).

The JPEG encoder can function down to a supply of 210mV, at a clock frequency of 5MHz (Fig. 27.3.3). Frequencies of 25, 50 or 100MHz are achieved at supplies of 300, 350 and 410mV, respectively. Fig. 27.3.4 shows a comprehensive state-of-the-art frequency comparison. The present work exceeds the speed performance of previous ultra-low-voltage designs in advanced nanometer technologies. Only Seok *et al.* [5] achieve a similar frequency, but their FFT core consumes 15.8nJ/transform, which is 550x higher than this work's minimum energy consumption. This large difference in energy consumption cannot be explained by the difference in computational work between an FFT and a JPEG encoder. Thus, given the similar performance figures of both designs, we conclude that the design presented here is much more energy efficient.

The minimum-energy point occurs at 330mV (Fig. 27.3.5), with the chip consuming 29.01pJ/pixel at an operating frequency of 41MHz. Overall, the JPEG encoder achieves an energy consumption of less than 50pJ/pixel for clock frequencies below 275MHz. The percentage of energy consumed by each block (as shown in Fig. 27.3.1) is also given. Observe that the register tables contribute significantly to leakage, as the quantization and zigzag and Huffman blocks have a much higher percentage of leakage than the 2D-DCT block, which does not contain a table. A total of 26 dies were measured. Across a supply range of 210 to 550mV, the mean variation (σ/μ) in operating frequency is 8.6% and the mean variation in energy consumption/pixel is 5.4%.

Figure 27.3.6 provides a state-of-the-art comparison between this work and another ultra-low-voltage JPEG encoder, fabricated in 65nm CMOS [6]. This work is able to function at a minimum supply of 210mV, while [6] is only able to reach a minimum of 400mV. The architecture of [6] consists of 4 parallel engines and a Huffman encoder that runs in a different voltage and clock domain, i.e. with a clock that is 4x the engine clock. At 400mV, the engines are able to operate at a clock frequency of 2.5MHz and the Huffman encoder runs at 10MHz at 600mV. This work achieves a frequency of 5MHz at the minimum supply, and 41MHz at the MEP, significantly outperforming [6]. Unfortunately, a direct comparison between the energy consumptions cannot be made, because [6] only provides the energy consumption per cycle, i.e. per pipeline stage. Since the number of pipeline stages is not mentioned in [6], it is not possible to calculate the total energy consumption, nor the energy-delay product (EDP).

This paper presents the design of a near-threshold JPEG encoder in 40nm CMOS that is able to function at ultra-low supply voltages as low as 210mV. The chip achieves significantly better than state-of-the-art operating frequencies, well within the MHz range, combined with low energy consumption. The variation resiliency of this JPEG encoder is validated by the low variation results of the measurements. We expect that the design principles used for this chip can be applied to other processor designs and will result in similar ultra-low-voltage characteristics.

References:

- [1] G.K. Wallace, "The JPEG Still Picture Compression Standard," *IEEE Trans. on Consumer Electronics*, vol. 38, no. 1, pp. xviii-xxiv, 1992.
- [2] M. Kovac and N. Ranganathan, "JAGUAR: A Fully Pipelined VLSI Architecture for JPEG Image Compression Standard," *Proceedings of the IEEE*, vol. 83, no. 2, pp. 247-258, 1995.
- [3] N. Reynders and W. Dehaene, "Variation-Resilient Building Blocks for Ultra-Low-Energy Sub-Threshold Design," *IEEE Trans. Circuits and Systems-II*, vol. 59, no. 12, pp. 898-902, 2012.
- [4] O. Weiss, M. Gansen and T.G. Noll, "A Flexible Datapath Generator for Physical Oriented Design," *European Solid-State Circuits Conf.*, pp. 393-396, 2001.
- [5] M. Seok, *et al.*, "A 0.27V 30MHz 17.7nJ/transform 1024-pt Complex FFT Core with Super-Pipelining," *ISSCC Dig. Tech. Papers*, pp. 342-343, 2011.
- [6] Y. Pu, *et al.*, "An Ultra-Low-Energy/Frame Multi-Standard JPEG Co-Processor in 65nm CMOS with Sub/Near-Threshold Power Supply," *ISSCC Dig. Tech. Papers*, pp. 146-147, 2009.

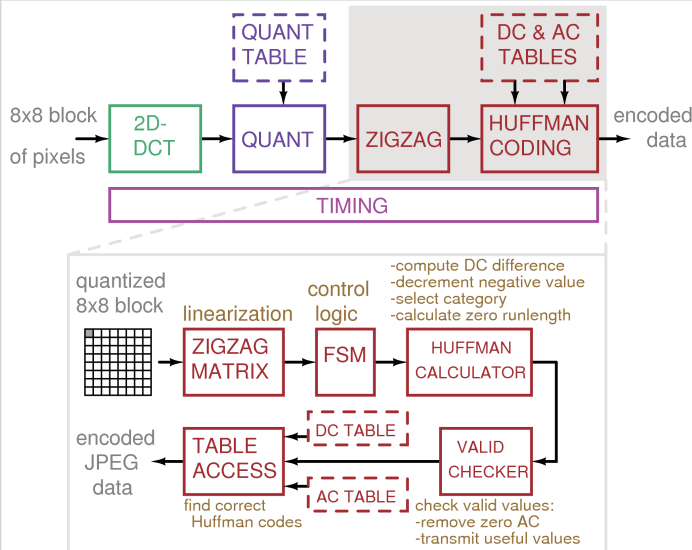


Figure 27.3.1: Block diagram of the JPEG encoder.

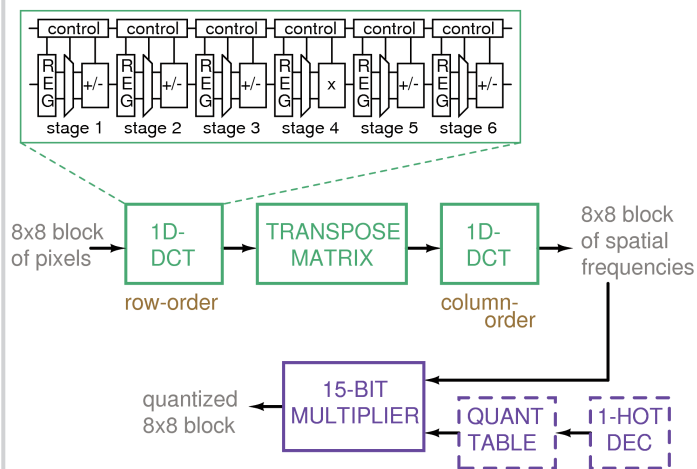


Figure 27.3.2: Implementation of the 2D-DCT and quantization.

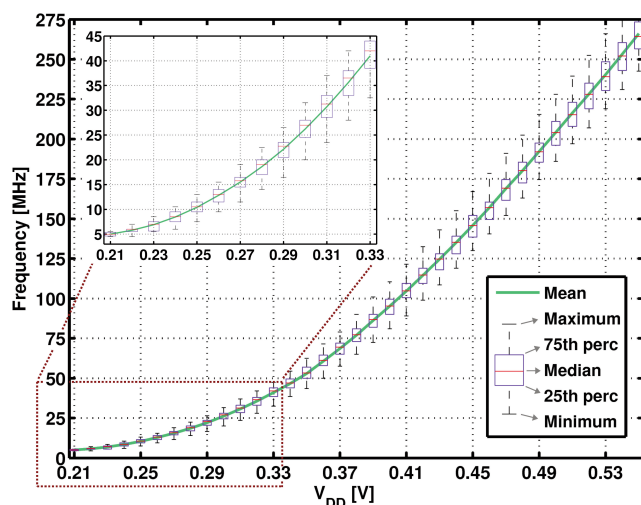


Figure 27.3.3: Boxplot of the measured maximum operating frequency as function of V_{DD} , obtained from 26 dies.

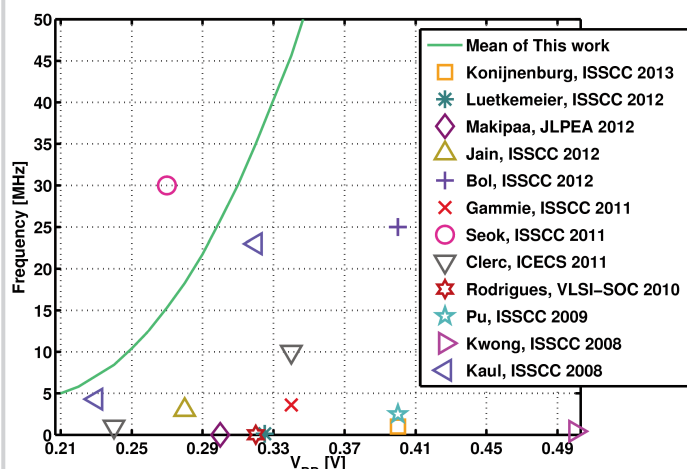


Figure 27.3.4: State-of-the-art frequency comparison between all other previously published ultra-low-voltage designs in advanced nanometer CMOS technologies.

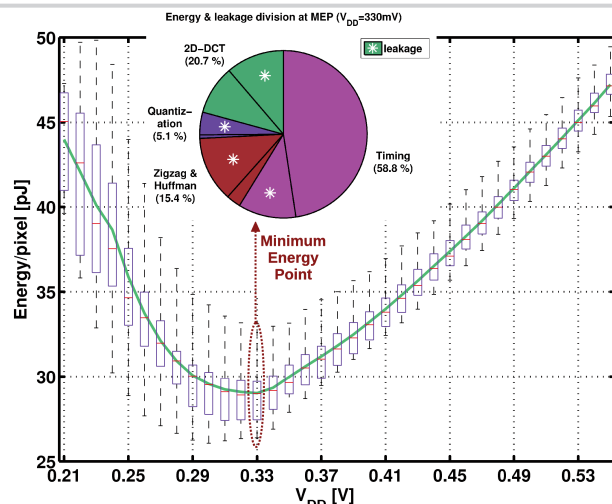


Figure 27.3.5: Boxplot of the measured energy consumption per operation as function of V_{DD} . The division of the energy consumption is also given at the MEP.

	This work	[6] ¹
CMOS Technology	40nm	65nm
Active area [mm ²]	0.557	1.960
Minimal voltage [V]	0.210	0.400 (engine) / 0.600 (Huff)
Voltage @ MEP [V]	0.330	0.400 / 0.600
Frequency @ MEP [MHz]	41.0	2.5 / 10
Energy/cycle @ MEP ² [pJ]	0.045	3 (for 4 engines) + 1.8
Energy/8x8 block @ MEP [pJ]	1857	-
EDP @ MEP [pJ. μ s]	45.29	-

¹ In [6], 2D-DCT and quantization are joined in a so-called engine and such an engine is operating at 1 voltage domain, while the Huffman encoder is operating in a second higher voltage domain.

² This is design-dependent, but it is the only energy figure provided in [6].

Figure 27.3.6: State-of-the-art comparison between ultra-low-voltage JPEG encoders in advanced nanometer CMOS technologies.

	Pipeline Depth
2D-DCT	458
Quantization	22
Zigzag	130
Huffman coding	36
Total JPEG	646

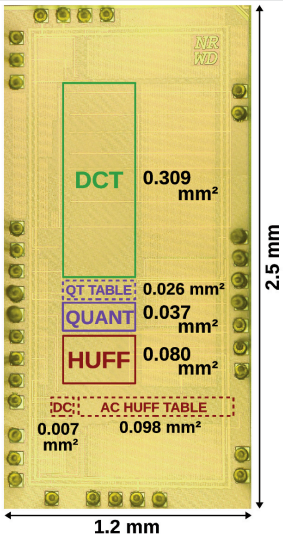


Figure 27.3.7: Chip micrograph of the 40nm JPEG encoder and pipeline depth overview. Active area is 0.557mm².